STIC Search Report
Biotech-Chem Library

STIC Database Tracking Number: 119538

TO: Changhwa Cheu **Location: rem/3c61** **Art Unit: 1641** **Thursday, April 15, 2004** **Case Serial Number: 09/807877**	**From: Barb O'Bryen** **Location: Biotech-Chem Library** **Remsen 1A69** **Phone: 571-272-2518** **barbara.obryen@uspto.gov**

Search Notes

[redacted]

PROTEIN SEQUENCES IN THE REGISTRY FILE

Please scan those reprints.

Sequence Selection

Protein sequence information in the Registry file is compiled by Chemical Abstracts Service (CAS) from novel sequences reported in original research articles, patent claims, and patent examples. CAS monitors several hundred original sources for sequence information, including major and minor serials, patents from 29 patent offices, and technical reports.

To date (3/93) there are over 216,000 protein sequences in the Registry file. Sequences with chain lengths of four or more are searchable by 1-letter or 3-letter amino acid codes. Dipeptides and tripeptides are also registered, but may be searched only by name or structure and not by sequence representation.

The Registry file contains protein sequences reported as early as 1957. These earlier records are mainly for smaller, naturally occurring peptides, such as angiotensins and insulins. The major growth of large protein sequence data has occurred since about 1980, coinciding with major breakthroughs in the DNA sequencing methods. Currently, approximately 2000 new protein sequences are added to the Registry file every month.

All newly reported sequences are registered and included in the Registry file. Entries are made for the first occurrence of a complete sequence and any major structure updates. A complete sequence is (a) one identified as complete by the author(s) or (b) one translated from a nucleic acid sequence, beginning with an AUG codon (or other author-identified start codon) and ending with a stop codon. Sequences of precursor forms (e.g., PRE-, PRO-, and PREPRO-) are registered separately from the mature form, when the cleavage sites are indicated.

The first occurrence of a sequence is determined by exact structure match of the reported structure with the existing records in the Registry file. Therefore, sequences that differ by even one in 350 amino acids are registered separately. Also, identical sequences with different chemical modifications, such as blocking groups, side chain substitutions, replacement with isotopes, are registered separately.

Sequences may be found in the Registry file for the following classes of proteins and peptides from both the journal and patent literature:

- Naturally occurring proteins and peptides
- Sequences deduced from gene translation and reported by the author
- Sequences deduced from the GenBank® database (gene translation)
- Chemically modified peptides and proteins
- Genetically engineered and synthetic proteins
- Multichain proteins
- Cyclic peptides
- Fusion proteins
- Peptide metal complexes
- Sequences containing uncommon amino acids, i.e. not genetically encoded

Partial sequences without internal gaps claimed in patents or related to the novelty of the claim have been indexed from 1988 onwards. Partial sequences containing 20 or more amino acids without internal gaps have been indexed from a select group of journals from 1991 onwards. Beginning in 1992, a partial sequence is registered from these journals for each protein for which there is at least one fragment of 20 or more contiguous and unambiguous amino acids. Protein sequences deduced from the nucleic acid sequences in the GenBank database are also available in the Registry file.

Sequence Family Search of Proteins (/sqsfp)

In the sequence family search, each amino acid in the query has to match either the exact amino acid or a family member equivalent, as shown in the Family Equivalence Table below. The Family Equivalence Table is applied only to each common amino acid in the sequence. Specific uncommon amino acids may be included in the sequence; however, family equivalents only exist for the common amino acids. An amino acid family is based on a conservative substitution of amino acids sharing a similar chemical property. Each common amino acid in the query is converted to its family class members in a search. A match occurs on a query sequence if each amino acid is exactly matched or any of its family members are encountered. For example, the Hydrophobic-Aromatic family consists of the common amino acids F, W, and Y. If the amino acid F is specified within a sequence exact family search, it will match on amino acids F, W, or Y.

FAMILY EQUIVALENCE TABLE

Family Class Name	**Family Class Members**
Neutral-Weakly Hydrophobic	Ala (A), Gly (G), Pro (P), Ser (S), Thr (T)
Hydrophilic-Acid Amine	Asn (N), Asp (D), Gln (Q), Glu (E)
Hydrophilic-Basic	Arg (R), His (H), Lys (K)
Hydrophobic	Ile (I), Met (M), Leu (L), Val (V)
Hydrophobic-Aromatic	Phe (F), Trp (W), Tyr (Y)
Crosslinking	Cys (C)

conservative substitutions allowed
in structure search & sequence search
in Registry File

=> fil reg; d stat que l12
FILE 'REGISTRY' ENTERED AT 14:48:09 ON 15 APR 2004
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.

Property values tagged with IC are from the ZIC/VINITI data file
provided by InfoChem.

STRUCTURE FILE UPDATES: 14 APR 2004 HIGHEST RN 675571-70-7
DICTIONARY FILE UPDATES: 14 APR 2004 HIGHEST RN 675571-70-7

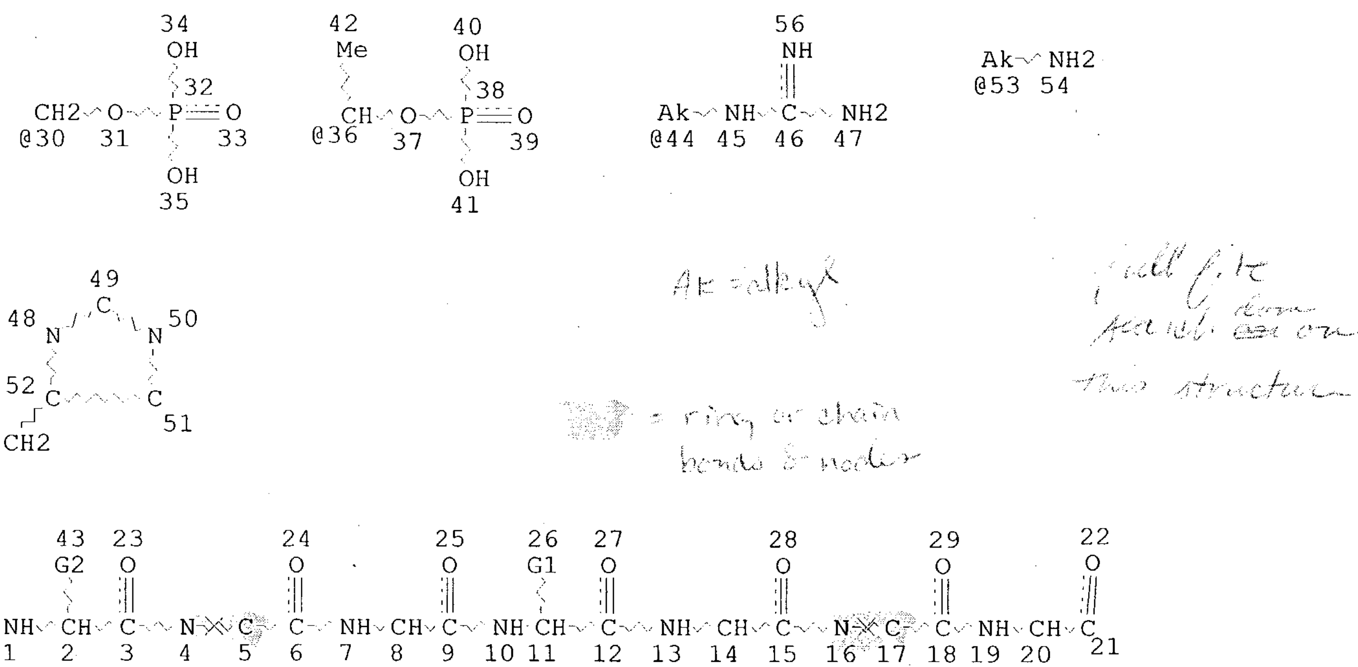
TSCA INFORMATION NOW CURRENT THROUGH JANUARY 6, 2004

Please note that search-term pricing does apply when
conducting SmartSELECT searches.

Crossover limits have been increased. See HELP CROSSOVER for details.

Experimental and calculated property data are now available. For more
information enter HELP PROP at an arrow prompt in the file or refer
to the file summary sheet on the web at:
http://www.cas.org/ONLINE/DBSS/registryss.html

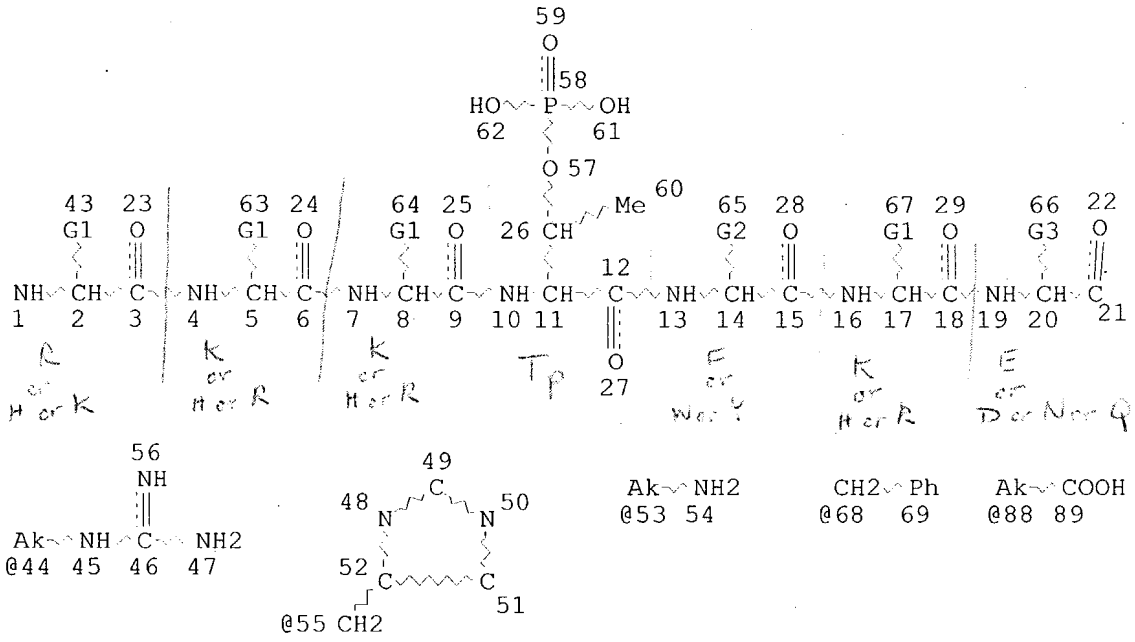
L5 STR



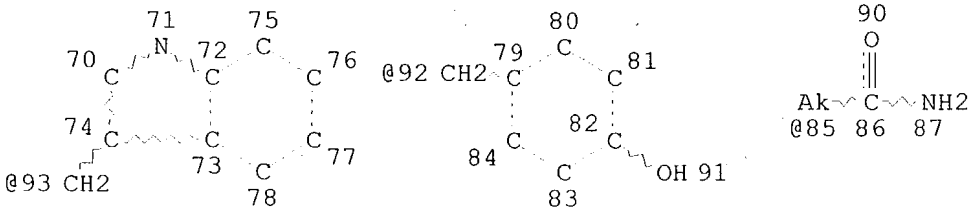
VAR G1=30/36
VAR G2=44/55/53
NODE ATTRIBUTES:
NSPEC IS RC AT 4
NSPEC IS RC AT 5
NSPEC IS RC AT 16
NSPEC IS RC AT 17
CONNECT IS E2 RC AT 44
CONNECT IS E2 RC AT 53
DEFAULT MLEVEL IS ATOM
DEFAULT ECLEVEL IS LIMITED

GRAPH ATTRIBUTES:
RING(S) ARE ISOLATED OR EMBEDDED
NUMBER OF NODES IS 56

STEREO ATTRIBUTES: NONE
L7 212 SEA FILE=REGISTRY SSS FUL L5
L8 STR



Page 1-A

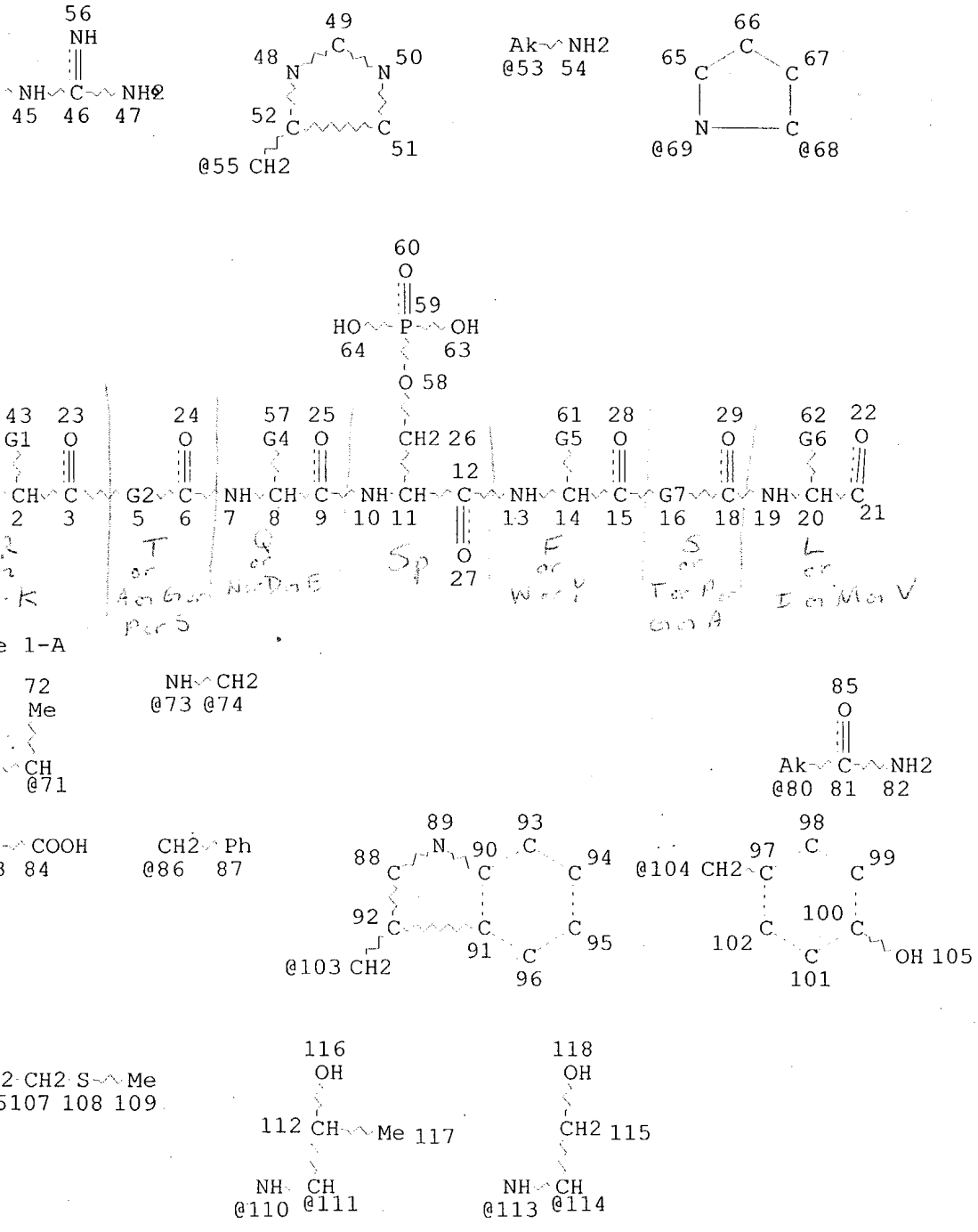


Subject search done looking for this structure or structure on next page

Page 2-A
VAR G1=44/55/53
VAR G2=68/93/92
VAR G3=85/88
NODE ATTRIBUTES:
CONNECT IS E2 RC AT 44
CONNECT IS E2 RC AT 53
CONNECT IS E2 RC AT 85
CONNECT IS E2 RC AT 88
DEFAULT MLEVEL IS ATOM
DEFAULT ECLEVEL IS LIMITED

GRAPH ATTRIBUTES:
RING(S) ARE ISOLATED OR EMBEDDED
NUMBER OF NODES IS 80

STEREO ATTRIBUTES: NONE
L10 STR



e 1-A

e 2-A
G1=44/55/53
G2=69-3 68-6/70-3 71-6/73-3 74-6/110-3 111-6/113-3 114-6
G4=80/83
G5=86/103/104
G6=I-BU/S-BU/106/I-PR
G7=69-15 68-18/70-15 71-18/73-15 74-18/110-15 111-18/113-15 114-18
C ATTRIBUTES:
NECT IS E2 RC AT 44
NECT IS E2 RC AT 53
NECT IS E2 RC AT 80

CONNECT IS E2 RC AT 83
DEFAULT MLEVEL IS ATOM
DEFAULT ECLEVEL IS LIMITED

GRAPH ATTRIBUTES:
RING(S) ARE ISOLATED OR EMBEDDED
NUMBER OF NODES IS 98

STEREO ATTRIBUTES: NONE
L12 0 SEA FILE=REGISTRY SUB=L7 SSS FUL (L8 OR L10)

100.0% PROCESSED 126 ITERATIONS 0 ANSWERS
SEARCH TIME: 00.00.01

=> fil hom
FILE 'HOME' ENTERED AT 14:48:12 ON 15 APR 2004

```
=> fil reg
FILE 'REGISTRY' ENTERED AT 15:06:34 ON 15 APR 2004
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
COPYRIGHT (C) 2004 American Chemical Society (ACS)

Property values tagged with IC are from the ZIC/VINITI data file
provided by InfoChem.

STRUCTURE FILE UPDATES:   14 APR 2004  HIGHEST RN 675571-70-7
DICTIONARY FILE UPDATES:  14 APR 2004  HIGHEST RN 675571-70-7

TSCA INFORMATION NOW CURRENT THROUGH JANUARY 6, 2004

  Please note that search-term pricing does apply when
  conducting SmartSELECT searches.

Crossover limits have been increased.  See HELP CROSSOVER for details.

Experimental and calculated property data are now available. For more
information enter HELP PROP at an arrow prompt in the file or refer
to the file summary sheet on the web at:
http://www.cas.org/ONLINE/DBSS/registryss.html
```

. = any amino acid

```
=> d que l22; fil hom
L16         53529 SEA FILE=REGISTRY ABB=ON  RTQ.FSL|RKK.FKE/SQSFP
L17         10266 SEA FILE=REGISTRY ABB=ON  (PHOSPH? OR PO2)/NTE
L20         65363 SEA FILE=REGISTRY ABB=ON  PHOSPHONO
L22             0 SEA FILE=REGISTRY ABB=ON  L16 AND (L17 OR L20)

FILE 'HOME' ENTERED AT 15:06:52 ON 15 APR 2004
```

- family search done allowing for conservative substitution